ODBITKA Z ROCZNIKÓW NAUK ROLNICZYCH
I LEŚNYCH

Stefan Moszczeński

Ujednostajnienie metod statystycznych dla opracowania materjałów rachunkowych z gospodarstw wiejskich

„Roczniki Nauk Rolniczych i Leśnych“ Tom XXII

Poznań 1929
Czcionkami Drukarni Uniwersytetu Poznańskiego

Autor z prośbą o [illegible]

ODBITKA Z ROCZNIKÓW NAUK ROLNICZYCH
I LEŚNYCH

Stefan Moszczeński

Ujednostajnienie metod statystycznych dla opracowania materjałów rachunkowych z gospodarstw wiejskich

„Roczniki Nauk Rolniczych i Leśnych" Tom XXII

Poznań 1929
Czcionkami Drukarni Uniwersytetu Poznańskiego



Stefan Moszczeński.

Ujednostajnienie metod statystycznych dla opracowania materjałów rachunkowych z gospodarstw wiejskich.

Z Zakładu Ekonomiki Gospodarstw Wiejskich Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie.]

(Wpłynęło dnia 9. III. 1929 r.)

Ustalenie mierników w rachunkowości rolniczej, a także liczb względnych, rozpatruje dr. W. Ponikowski w referacie p. t.: „Ustalenie niektórych mierników i liczb względnych w rachunkowości rolniczej". To zagadnienie ma niesłychanie wielkie znaczenie dla samej rachunkowości i dla przedstawienia jej wyników na międzynarodowym terenie rolnictwa. Jeśli tak jest, a że tak jest, chyba nikt w to nie wątpi, to nie mniej ważnem, bo właściwie uwieńczeniem samej pracy, jest ujednostajnienie metod statystycznych, za pomocą których badacze ekonomiki gospodarstwa wiejskiego mają opracowywać porównawcze dane, których dostarcza rachunkowość.

Zastosowanie statystyki w najprostszej formie dla scharakteryzowania jakiejś cechy zjawisk ekonomicznych za pomocą średniej arytmetycznej nie jest nową rzeczą. Już w pierwszej połowie XIX w. interesowano się miarami przeciętnemi przy porównywaniu ilości i wartości poszczególnych składników gospodarskich, a nawet niektórych kosztów. Jednak zastosowanie metod statystycznych do badania funkcji zachodzących między zjawiskami gospodarczemi datuje się w nauce rolnictwa dopiero od początku obecnego stulecia.

W tym czasie, kiedy badacze ekonomiki gospodarstwa wiejskiego uświadamiali sobie olbrzymią rolę, jaką mogą odegrać porównania statystyczne w rozwoju ich dyscypliny naukowej, teorja statystyki była już dawno znana. Wiele już metod było

opracowanych w innych dziedzinach wiedzy, i to zarówno metod statystyki prostej, niematematycznej, jak i t. zw. matematycznej, t. j. opartej na rachunku prawdopodobieństwa.

Około roku 1896 jeden z pierwszych, Rodewald z Kielu, użył błędu prawdopodobnego, jako miary dokładności w doświadczeniach nad kiełkowaniem, w doświadczeniach polowych, a także w ocenie zbiorów ziarna z omłotów próbnych, a więc również w zabiegach natury ekonomicznej. Rodewald, mówiąc o zastosowaniu rachunku prawdopodobieństwa i o błędzie prawdopodobnym, jako o metodzie oceny, dodał: „Auch dürfte die landwirtschaftliche Betriebslehre ein fruchtbares Feld der Anwendung dieser Methode bieten".

Myśl rzuconą przez Rodewalda podjął A. Mitscherlich. Już w roku 1903 ukazuje się praca: „Die Schwankungen der landwirtschaftlichen Reinerträge berechnet für einige Fruchtfolgen mit Hilfe der Fehlerwahrscheinlichkeitsrechnung", w której to pracy Mitscherlich posługuje się do oceny materjałów statystycznych błędem prawdopodobnym średniej arytmetycznej (E_A).

Mitscherlich badał zbiory i ceny przeciętne zbóż i ziemniaków w Niemczech, koszty produkcji roślin, dochody czyste, czas niezbędny do wykonywania poszczególnych robót, ilość dni sprzężajnych, zużytych średnio w ciągu roku i t. p. zjawiska gospodarcze. Następnie z materjałów zebranych obliczał błędy prawdopodobne i kreślił linje częstotliwości empiryczne i teoretyczne, to jest takie, jakie istniałyby, gdyby zbiorowisko zdarzeń rozpatrywanych obejmowało wszystkie zdarzenia danego rodzaju. Z zestawień obu linji można było wnioskować, czy odchylenia od średniej mogą być uważane za błędy przypadkowe, czy też oznaczają, że zaszły jakieś okoliczności znamienne. Prócz tego Mitscherlich badał za pomocą spółczynnika zmienności $\left(\frac{A}{E}\cdot 100\right)$ wiarogodność materjałów statystycznych, choć metoda, którą obrał dla tego celu, nie była trafnie pomyślana. Stosunek błędu średniego, czy prawdopodobnego do średniej arytmetycznej jest wyrazem dokładności pomiarów tylko przy pomiarach tej samej wielkości fizycznej, użyty zaś do charakteryzowania zbiorowiska różnych osobników, staje się odzwierciedleniem bogactwa wahań.

Już w rok po Mitscherlich'u Stefan Pawlik, prof. z Dublan pod Lwowem, zastosował błąd prawdopodobny jako miarę stałości plonów chmielu w Starym Siole w Małopolsce. Błędem prawdopodobnym (E) posługują się w późniejszych czasach niejednokrotnie autorzy niemieccy w nauce ekonomiki gosp. wiejsk. Należy zaznaczyć, że do charakterystyki zjawisk ekonomicznych ta miara nie zawsze się nadaje. Błąd prawdopodobny jest tylko wtedy odpowiednią miarą, jeżeli zjawiska badane ulegają prawu Gauss'a.

Wogóle zjawiska ekonomiczne są jeszcze mało poznanym terenem badań statystycznych. Nie wiemy, jakie to są zjawiska ekonomiczne, które ulegają prawu Gauss'a. Z góry jesteśmy skłonni przypuszczać, że to prawo nie często urzeczywistnia się wśród nich. Wprawdzie na pojawianie się zdarzeń ekonomicznych oddziaływują bardzo liczne, może nawet niezliczone przyczyny, ale wiele z nich jest istotnych, a nie przypadkowych, a powtóre między zdarzeniami ekonomicznemi często istnieje widoczna zależność, co również wyklucza rozsianie normalne wśród badanych zjawisk. Czy to rozpatrujemy płace, czy ceny na rynkach, łatwo się domyślamy, że płace w jednym okręgu są zależne od płac w drugim okręgu, ceny w jednym okręgu kształtują się pod wpływem cen w innych okręgach i t. p.

Chcąc stosować krzywą Gauss'a, musielibyśmy każdorazowo wpierw zbadać za pomocą funkcyj krytycznych, czy wyniki pomiarów odpowiadają założeniom krzywej tego typu. Zbyteczny to trud. Niektórzy autorzy posługują się błędem średnim, nawet w tych dziedzinach badań, w których z natury rzeczy stosowanie błędu prawdopodobnego ma z góry pewne uzasadnienie, a to wychodząc z założenia, że jeśli są spełnione warunki niezbędne dla błędu prawdopodobnego, to tem samem są spełnione dla błędu średniego. My tem bardziej w naszych badaniach nad wynikami rachunkowości rolniczej będziemy posługiwać się tylko odchyleniem średniem (σ).

Niezależnie, czy badane zjawiska przez Mitscherlich'a, Pawlika, Pauli'ego i innych były właściwie charakteryzowane błędem prawdopodobnym, czy nie lepiej byłoby posłużyć się błędem średnim, stwierdzić musimy, że metoda ta nie dała ekonomice pozytywnych wyników, a przez to zniechęciła badaczy

do dalszych prób z zastosowaniem statystyki matematycznej. Omyłka wspomnianych badaczy polegała na tem, że im zależało głównie na scharakteryzowaniu bogactwa wahań, co wprawdzie w analizie zjawisk z innych dziedzin życia ma nieraz pierwszorzędne znaczenie, ale w ekonomice gospodarstwa wiejskiego schodzi na dalszy plan.

Prof. Laur i prof. Waterstradt, obaj pierwszorzędni przedstawiciele ekonomiki gospodarstwa wiejskiego, pierwszy w Szwajcarji, drugi w Niemczech, próbowali także posługiwać się błędem prawdopodobnym dla scharakteryzowania bogactwa wahań w zjawiskach przez siebie badanych, lecz nie widząc praktycznych wyników, rychło poszli własną drogą. Przedewszystkiem zarzucili metodę charakteryzowania spostrzeganych wyników miarami rozsiania, a więc zarzucili metodę statystyki matematycznej, powtóre zainteresowali się zgoła innem zagadnieniem, a mianowicie odkrywaniem związków funkcyjnych między zjawiskami gospodarczemi, i to w możliwie najprostszy sposób.

Ten prosty sposób wprowadzony przez prof. Laura[1]) i prof. Waterstradta[2]) do ekonomiki gospodarstwa wiejskiego, polega na ułożeniu dwóch lub więcej szeregów liczb zmiennych, z których każdy charakteryzuje inną cechę. Jeden szereg jest uporządkowany według wielkości rosnących lub malejących jednej cechy badanych osobników (niech osobnikiem będzie jakaś majętność), a w drugim szeregu, ewentualnie w następnych, układa się wymiary cech według przynależności do osobników w pierwszym szeregu. Dla uzyskania większej przejrzystości w szeregach, gdy pomiarów jest wiele, tworzymy z wymiarów klasy, tak że cechy badane porównywamy klasami.

Tą samą metodą posługiwałem się w mej pracy w roku 1913[3]).

Przytaczam z tej pracy parę przykładów.

[1]) „Untersuchungen betreffend die Rentabilität der schweiz. Landwirtschaft im Erntejahr 1907". Bern 1909.

[2]) Tühnen-Archiv" 1906.

[3]) Stefan Moszczeński: Gdzież są granice intensywności?

Pierwszy przykład.

Wpływ odległości majątku od stacji dróg żelaznych na dochodowość.

Klasy	Majątki leżące od stacji drogi żelaznej	Liczba majątków w każdej klasie	Średni dochód czysty z morga z majątków w każdej klasie
I	bliżej niż 15 wiorst	15	16,06 rb.
II	15—30 „	8	11,48 „
II	30—45 „	3	7,71 „
IV	dalej niż 45 „	1	5,90 „

Spostrzegamy wyraźnie, jak w miarę zwiększania się odległości od stacji dróg żelaznych dochodowość majątków spada.

Drugi przykład.

Wpływ rozmiaru uprawy roślin pastewnych na dochodowość gospodarstwa

Rozmiar uprawy roślin pastewnych w stosunku do ogólnego areału gruntów ornych	Liczba majątków	Średni dochód czysty z morga
mniej niż 10%	7	20,69 rb.
10—20 „	24	14,02 „
20—30 „	6	13,89 „
ponad 30 „	1	5,63 „

Spostrzegamy wyraźnie, jak w miarę zwiększania się rozmiaru uprawy roślin pastewnych dochody czyste maleją.

Trzeci przykład.

Wpływ rozmiaru uprawy roślin okopowych na dochodowość gospodarstwa.

Rozmiar uprawy roślin okopowych w stosunku do ogólnego areału gruntów ornych	Liczba majątków	Średni dochód czysty z morga
mniej niż 10%	2	9,23 rb.
10—20 „	11	11,77 „
20—30 „	17	15,10 „
ponad „	4	24,32 „

Spostrzegamy, jak w miarę zwiększania się uprawy roślin okopowych dochody czyste podnoszą się.

W tymże samym roku t. j. 1913 dokonano podobnych badań w Stanach Zjednoczonych Ameryki Północnej w stanie Yowa na bardzo licznym materjale. Oto dwa przykłady:

Wpływ rozmiaru pastwiska na dochodowość 965 ferm w 1913 r.

% ziemi pod pastwiskiem	Liczba ferm	Dochód z pracy w $
poniżej 10	68	485
11—20	303	429
21—30	363	307
31—40	160	180
powyżej 40	71	85

W tym samym stanie Yowa dokonano badań w roku 1918 a więc u schyłku wielkiej wojny.

Wpływ rozmiaru pastwiska na dochodowość 210 ferm w 1918 r.

% ziemi pod pastwiskiem	Liczba ferm	Dochód z pracy w $	Rozmiar ferm w akrach
poniżej 10	16	2 188	159
11—20	36	1 876	202
21—30	78	1 506	212
31—40	56	914	233
powyżej 40	24	789	282

W ten prosty sposób badamy współzależność między różnemi zabiegami rolnika a dochodem czystym z gospodarstwa. Niezmiernie ciekawe są badania wpływu poszczególnych cech intensywności, czy to rozszerzonej uprawy okopowych, czy intensywniejszego nawożenia, czy meljoracji, czy gałęzi przemysłowych i t. p. Gdybyśmy mogli dostatecznie trafnie scharakteryzować wszystkie cechy intensywności jednym wspólnym miernikiem, do czego zmierza wspomniany wyżej referat dr. Wacława Ponikowskiego, oświetlilibyśmy rolnikowi ogólną tendencję

opłacalności na tem niezmiernie ciekawem polu jego pracy. Rozumiemy dobrze, że nie należy zacieśniać się w zestawianiu cech z dochodami czystemi. Również koszty produkcji i dochody brutto są ważną zmienną zależną od wielu czynników działających w gospodarstwie.

II.

Obecnie w literaturze europejskiej ekonomicznej spotykamy przeważnie proste charakterystyki liczbowe za pomocą średniej arytmetycznej (*A*) i badanie współzależności za pomocą prostego zestawienia dwóch lub więcej szeregów zmiennych. Można jednak opracowywać uzyskane liczby za pomocą wzorów statystyki matematycznej. Różnica, jaka istnieje między obu metodami, nie zasadza się bynajmniej tylko na trudnościach rachunkowych. Statystyka matematyczna opiera się na rachunku prawdopodobieństwa, który odpowiada ściślej zmienności życia, aniżeli proste reguły statystyki niematematycznej.

Poglądy wielu ekonomistów, że życie gospodarcze nie układa się według reguł matematycznych, lecz że do jego poznania trzeba iść po drodze tylko intuicji i zdrowego sensu, nie są dostatecznie uzasadnione. Wśród zjawisk ekonomicznych istnieje pewna stałość faktów. Szereg zdarzeń powtarza się z roku na rok, lub z jakiegoś okresu na okres z zadziwiającą prawidłowością. Wspomnę tu choćby o wypadkach losowych, jakiemi są pożary. Bez istnienia stałości wypadków instytucje ubezpieczeniowe nie rozwijałyby się prawidłowo. Na tej samej podstawie stałości zdarzeń oparte są budżety państwowe, komunalne, prywatne. Praca rolnika musi być rozpatrywana pod takim samym kątem. Jego coroczny preliminarz robót, przychodów i rozchodów, naturalji i gotówki, jego planowanie organizacji i zarządzeń poszczególnych, nawet metody uprawy i chowu opierają się na prawidłowościach, które istnieją w przyrodzie i w życiu gospodarczem, a których poznanie określa on sam słowem: doświadczenie.

Rolnik nie przypuszcza, jak często, jak nieomal na każdym kroku, posługuje się prawdopodobieństwem statystycznem. Rozumie się, robi to intuicyjnie i kieruje się przytem zdrowym

sensem. Nie są to naogół źli doradcy, jeśli intuicja jest trafna i głęboka, a sens jest rzeczywiście zdrowy. Istnieje nawet ścisła łączność między temi doradcami a rachunkiem prawdopodobieństwa. Przecież rachunek prawdopodobieństwa, według słów Laplace'a nie jest niczem innem, jak zdrowym sensem przeożonym na rachunek.

Ten zdrowy sens mówi nam, że średnia arytmetyczna, nawet trafnie zastosowana jako miara położenia, charakteryzuje ściśle daną zbiorowość tylko w tym wypadku, kiedy wyczerpane są wszystkie możliwe spostrzeżenia.

Zliczymy n. p. we wszystkich gospodarstwach badanych wszystką ilość koni lub bydła, wszystkich ordynarjuszów, ilość mieszkań, powierzchnię wszystkich użytków i t. p., w takim razie średnie arytmetyczne obliczane na podstawie tego pełnego materjału są rzeczywiście średniemi arytmetycznemi zbiorowości badanej. Ale dokonywanie pomiarów wszystkich cech, a cech tych jest wiele, przechodzi siły nasze, np. oznaczenie we wszystkich gospodarstwach obrotów kasowych, dochodów czystych, dochodów brutto, kosztów produkcji, dni pracy i t. p. Wtedy zadawalniamy się mniejszą ilością spostrzeżeń i z nich przenosimy wnioski na całą zbiorowość.

Ilość gospodarstw, z których możemy mieć materjały rachunkowe do rozporządzenia, stanowi nieraz tylko ułamek procentu ogółu gospodarstw, o które nam chodzi. Tak np. w Polsce na 3 200 000 gospodarstw chłopskich zamyka się mniej więcej rocznie rachunki z 1 500 gospodarstw, co stanowi zaledwie około 0,05%. Z tej tak bardzo nielicznej garstki gospodarstw t. j. z tak zwanej zbiorowości próbnej przenosimy otrzymaną charakterystykę liczbową na wszystkie gospodarstwa chłopskie w Polsce t. j. na zbiorowość generalną. Rozumie się, nasza średnia arytmetyczna jest tylko przybliżeniem średniej arytmetycznej jakiejś cechy zbiorowości generalnej. Przenosząc średnią arytmetyczną z naszej zbiorowości próbnej na zbiorowość generalną, popełniamy błąd. Jak wielki to jest błąd, nie dowiemy się z wzorów statystyki niematematycznej, a jednak jest rzeczą ważną poznać go, bo tem samem poznajemy, w jakiem przybliżeniu nasza średnia może być uważana za przedstawicielkę całej zbiorowości. Statystyka

matematyczna daje nam łatwy do obliczenia wzór błędu średniego średniej arytmetycznej:

$$\sigma_A = \frac{\sigma}{\sqrt{n-1}}$$

gdzie σ_A oznacza błąd średni średniej arytmetycznej, σ odchylenie średnie, względnie błąd średni, n ilość dokonanych oznaczeń.

Błąd średni średniej posiada tę właściwość, że wyznacza prawdopodobne granice, w których mieści się prawdziwa wartość średniej, t. j. średniej całej zbiorowości generalnej. Umieszczamy ten błąd obok średniej w następujący sposób:

$$A \pm \sigma_A$$

Innemi słowy, ten wzór mówi, jak wielki popełniamy błąd (in + lub in —), kiedy chcemy wnioskować ze średniej arytmetycznej próby o średniej arytmetycznej zbiorowości generalnej. Im błąd w stosunku do średniej większy, tem z mniejszą pewnością przenosimy wyniki ze zdarzeń zauważonych na niezauważone.

Wzór błędu średniego średniej arytmetycznej wyżej podany

$$\frac{\sigma}{\sqrt{n-1}}$$

jest ścisły, jeżeli się odnosi do pomiarów tej samej wielkości fizycznej. Można go uważać praktycznie za wystarczająco ścisły, jeżeli zbiorowość generalna jest bardzo liczna, np. zbiorowisko kwiatów tej samej odmiany, ryb w morzu, czy coś podobnego oczywiście przy zachowaniu wszystkich po temu potrzebnych warunków. Wielkość tego błędu zależy bowiem od stosunku ilościowego zdarzeń zbiorowości próbnej do zdarzeń zbiorowości generalnej. Im więcej liczebność zbiorowości próbnej zbliża się do liczebności zbiorowości generalnej, tem z większą pewnością, a więc z mniejszym błędem, możemy średnią próby zastosować do zbiorowości generalnej. Jeśli oznaczenia zbiorowości próbnej wyczerpały całą zbiorowość generalną, niema mowy o jakimkolwiek

błędzie średniej, bez względu na wielkość odchylenia średniego poszczególnych wyników pomiaru.

Wzór $\frac{\sigma}{\sqrt{n-1}}$ nie nadaje się do ujmowania tej logicznie nasuwającej się zależności błędu średniego średniej od stosunku oznaczeń w obu zbiorowiskach. Dr. Jerzy Spława-Neyman podaje w swej pracy[4]) następujący wzór błędu średniego średniej arytmetycznej:

$$\sigma_A = \sqrt{\frac{N-n}{N(n-1)}} \cdot \sigma,$$

gdzie N oznacza liczebność zbiorowości generalnej, n liczebność zbiorowości próbnej. Widać z wzoru, że jeśli n zbliża się swą wielkością do N, błąd średni średniej szybko maleje. Jeśli $n = N$, to znaczy zostały wyczerpane wszystkie możliwe oznaczenia, spółczynnik przy σ równa się zeru, a wtedy błąd średni średniej także równa się zeru.

W ekonomice gospodarstw wiejskich osobnikami, których cechy są badane, bywają często warsztaty rolne. Ilość warsztatów rolnych na danym terenie jest przeważnie znana, mamy więc łatwo do czynienia ze skończoną ilością pomiarów, a więc tem samem w opracowaniach statystycznych gospodarstw wiejskich powinniśmy się posługiwać w wielu wypadkach wzorem błędu średniego średniej arytmetycznej:

$$\sigma_A = \sqrt{\frac{N-n}{N(n-1)}} \cdot \sigma.$$

Z dotychczasowych uwag wynika, że należy każdą średnią arytmetyczną zbiorowości próbnej traktować z zastrzeżeniem, t. j. podawać przy niej jej błąd średni. Jeszcze z większą ostrożnością należy postępować, kiedy przystępujemy do porównywania między sobą średnich arytmetycznych dwóch zbiorowości próbnych.

[4]) „Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych". Roczniki Nauk Rolniczych, Tom X, Zeszyt 1. Poznań 1923 r.

Różnica między temi średniemi charakteryzuje obie zbiorowości próbne, ale ponieważ średnie arytmetyczne zbiorowości próbnych różnią się od średnich arytmetycznych zbiorowości generalnych, nie możemy wnioskować, że różnica znaleziona między zdarzeniami zauważonemi odnosi się do wszystkich zdarzeń.

By się przekonać, czy istnieje słuszna podstawa do mniemania, że otrzymana różnica między średniemi charakteryzuje trafnie obie zbiorowości generalne, posługujemy się wzorem błędu średniego różnicy dwóch średnich arytmetycznych:

$$\sigma_{A_1-A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2}}$$

jeśli obie zmienne są całkowicie niezależne, a w wypadku zależności między zmiennemi wzorem:

$$\sigma_{A_1-A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2} - 2 \cdot r \cdot \sigma_{A_1} \cdot \sigma_{A_2}}$$

gdzie $\sigma_{A_1-A_2}$ oznacza błąd średni różnicy dwóch średnich arytmetycznych A_1 i A_2, σ_{A_1} i σ_{A_2} błędy średnie średnich arytmetycznych, r spółczynnik współzależności między obiema zmiennemi.

Jak już wspominaliśmy, wypadki współzależności między zmiennemi występują dość często w zjawiskach ekonomicznych, np. kiedy badamy różnice poziomu cen na różnych rynkach, płac w różnych dzielnicach lub państwach i t. p. a więc zastosowanie wzoru:

$$\sigma_{A_1-A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2} - 2\, r \cdot \sigma_{A_1} \cdot \sigma_{A_2}}$$

będzie często właściwą rzeczą.

Uważamy różnicę między dwoma średniemi arytmetycznemi zbiorowości próbnej za istotną, tj. charakteryzującą zbiorowości generalne, jeśli ta różnica jest większa od potrójnego błędu średniego różnicy. Im ta różnica między średniemi jest większa w stosunku do potrójnego błędu różnic, tem z większem prawdopodobieństwem możemy ją przenieść na zbiorowość generalną.

Gdybyśmy dokonali wszystkich możliwych oznaczeń w obu zbiorowościach, nie istniałby ani jeden z błędów średniej, a tem samem i błąd średni różnic. Różnice między średniemi obu zbiorowości byłyby w takim wypadku ścisłym wyrazem różnic ich wielkości.

Wynalezienie metody oceniania różnic między średniemi jest wielką zdobyczą statystyki teoretycznej. Metoda ta powstrzymuje od zbyt pochopnego wnioskowania. Różnice między średniemi mogą być znaczne, a mimo to nie będą znamienne, jeśli błędy są duże.

Obliczanie przeto zarówno błędu średniego średniej, jak i błędu średniego różnic między średniemi, jest nieodzownem zadaniem, jeśli chcemy nadać średniej arytmetycznej właściwe znaczenie. Nie od rzeczy będzie tu wspomnieć, że średnia arytmetyczna, choćby była podana z swym błędem, stać się może liczbą bez głębszej treści, jeśli nie zachowane są pewne warunki badania. Niech np. średnia arytmetyczna charakteryzuje jakąś cechę gospodarstw, leżących w warunkach ekonomicznych zbyt niepodobnych do siebie, albo gospodarstw zbyt różnego typu, zbyt różniących się wielkością, albo niech średnia przedstawia poziom cen osobników o zbyt różnej wartości użytkowej, to taka średnia będzie pozbawiona sensu. W takich wypadkach właściwiej będzie posłużyć się inną miarą położenia, czy to wartością najczęstszą, czy średnią geometryczną, czy może nawet wartością środkową.

III.

Drugie zagadnienie, które należałoby ustalić na terenie międzynarodowych badań statystycznych, dotyczy ujmowania współzależności między zjawiskami w gospodarstwach wiejskich. Podawaliśmy powyżej, w jak prosty sposób przedstawia się ową współzależność, a mianowicie za pomocą odpowiedniego ułożenia dwóch lub więcej szeregów liczb zmiennych. Metoda ta daje dużą dowolność badaczowi, który może oświetlać dane zagadnienie w sposób, jaki zechce. Na dowód powyższego przytoczę następujące przykłady. Mam przed sobą materjał

dotyczący rozmiaru uprawy buraka cukrowego i dochodów czystych z 33 gospodarstw ze Śląska z czasów przedwojennych. Dane podaję uporządkowane w dwa szeregi.

Powierzchnia pod burakami w 0/0 gruntu ornego	Dochód czysty w mk z ha	Powierzchnia pod burakami w 0/0 gruntu ornego	Dochód czysty w mk z ha
0	— 19,3	7,3	80,0
0	19,5	8,0	146,7
0	19,5	8,8	56,0
0	33,7	9,5	41,6
0	55,5	10,0	— 9,1
0	72,0	11,0	79,5
0	72,4	12,5	17,2
0,4	— 23,9	14,0	142,0
1,8	121,3	14,2	162,4
3,0	84,3	14,2	8,0
4,0	48,5	15,6	73,2
4,5	41,1	16,0	224,9
5,6	53,6	17,0	99,8
5,9	123,4	18,0	96,5
6,0	92,1	20,0	— 12,3
6,8	28,8	20,0	147,4
		23,0	98,3

Badacze mogą ująć ten materjał liczbowy w klasy w dowolny sposób. Załóżmy dwóch badaczy. Obaj wprowadzają podział na klasy nieprawidłowo, jak to się dotychczas zwykle robi w literaturze ekonomicznej (p. prof. Laur i inni uczeni).

Badacz I.

Powierzchnia pod burakami w °/₀ gruntów ornych	Liczba gospodarstw w klasie	Dochód czysty w mk z ha średnio
bez buraków	7	36,2
do 5°/₀	5	54,3
5—10°/₀	9	68,1
10—15°/₀	5	81,8
ponad 15°/₀	7	103,9

Badacz II.

Powierzchnia pod burakami w % gruntów ornych	Liczba gospodarstw w klasie	Dochód czysty w mk z ha średnio
bez buraków	7	36,2
do 5%	5	54,3
5—10%	9	68,1
10—18%	9	100,6
ponad 18%	3	77,8

Pierwszy badacz będzie dowodził, że według danych zebranych przed wojną nie znaleziono granicy opłacalności. Im więcej się uprawiało buraków w stosunku do powierzchni gruntów ornych, tem większe były dochody czyste. Drugi badacz powie, że współzależność istnieje, ale jest krzywolinijna. Uprawa buraków cukrowych opłacała się do pewnych granic, nie wyżej 18% powierzchni gruntów ornych. Jeżeli powierzchnia pod burakami przekraczała te granice, dochody, średnio biorąc, spadały.

Celem otrzymania więcej objektywnej miary posługujemy się wzorami statystyki matematycznej. Spółczynnik współzależności (r) wykazuje zależność istniejącą między dwoma zmiennemi. Ktokolwiek obliczy spółczynnik między jakiemikolwiek dwoma zmiennemi, otrzyma zawsze tę samą wielkość. Drugą dodatnią cechą spółczynnika współzależności jest to, że się go wyraża jedną liczbą. Możemy przeto w łatwy sposób stosunki zależności między zmiennemi zestawiać i porównywać. Interesowałoby nas między innemi, w jakim okresie czasu współzależność między rozmiarem pastwiska a dochodowością ferm była większa w stanie Yowa, czy w roku 1913, czy w roku 1918. Tej różnicy nam nie wyrazi zestawienie dwóch szeregów zmiennych. Jedynie za pomocą spółczynnika współzależności możemy stosunek wyrazić.

W naszym przykładzie z burakami spółczynnik współzależności $r = 0{,}34$. Jest to niewielki spółczynnik. Wnioskujemy z niego, że współzależność między uprawą buraków a dochodami czystemi jest słaba w rozpatrywanych majątkach, chociaż tablica, którą ułożył badacz I, wskazuje na bardzo wyraźną zależność.

Idąc za wskazaniami badacza I, dążylibyśmy do rozszerzenia uprawy buraka. Kierując się wynikiem, otrzymanym z zastosowania wzoru spółczynnika, będziemy ostrożni, i to podwójnie ostrożni. Po pierwsze, spółczynnik wskazuje na małą współzależność, powtóre każdy wzór statystyki matematycznej nasuwa nam pytanie, w jakim stopniu charakterystyka zbiorowości próbnej jest dobrem odzwierciedleniem zbiorowości generalnej.

Wzięliśmy próbę z 33 majątków. Jest to maleńka cząstka wszystkich majątków w danym okręgu. Chcąc zbadać, w jakich granicach odchyla się współzależność między zmiennemi generalnemi od współzależności zmiennych próbnych, posługujemy się błędem średnim spółczynnika współzależności (σ_r).

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}$$

Błąd średni spółczynnika r oznacza wielkość błędu, jaki popełniamy, kiedy chcemy przenieść znalezioną współzależność na wszystkie możliwe oznaczenia, tak samo jak błąd średni średniej wyznaczał granice, w których mieści się prawdopodobnie średnia zbiorowości generalnej. Im błąd większy przy tym samym spółczynniku, czy przy tej samej średniej, tem większych stosunkowo wahań należy się spodziewać. Oczywistą jest rzeczą, że przy dużych wahaniach zmniejsza się prawdopodobieństwo trafnego scharakteryzowania zbiorowości generalnej, a liczby charakterystyczne, czy to będzie średnia, czy spółczynnik współzależności tracą sens.

Przyjęto za regułę, że spółczynnik współzależności między dwoma zbiorowościami, z których tylko część oznaczeń bierzemy pod uwagę, ma tylko wtedy znaczenie, jeśli jest trzy razy większy od swego błędu średniego. Ponieważ przy ugrupowaniach, nie stosujących się do prawa Gauss'a, wzór $\frac{1 - r^2}{\sqrt{n}}$ nie jest ścisły, przeto poleca się dla pewności przyjmować za obszar krytyczny obszar znacznie przewyższający potrójny błąd średni. Z tego wynikałoby, że spółczynnik r mniejszy od potrójnej wielkości swego błędu, choćby był bardzo duży, nie jest wart rozważania, kiedy go zechcemy przenieść na zbiorowość całkowitą.

W naszym wypadku błąd średni wynosi 0,15, a więc

$$r = 0{,}34 \pm 0{,}15.$$

Ponieważ w naszym przykładzie spółczynnik jest zaledwie przeszło dwa razy większy od swego błędu, przeto nie mamy prawa stawiać żadnych wniosków co do stosunków, panujących wśród majątków na całym badanym terenie. Ten wynik jest zgodny ze zdrowym sensem. Jeśli bowiem jest niewiele spostrzeżeń, a przytem spostrzeżenia, jak w naszym przykładzie, wykazują duże wahania, to z takiego materjału nie można wyciągać żadnych wniosków.

Gdyby jednak liczba spotrzeżeń wynosiła 400, a współzależność między uprawą buraków a dochodami czystemi w zbiorowości próbnej wyrażała się tym samym spółczynnikiem $r = 0{,}34$, w takim razie błąd średni spółczynnika równałby się 0,044, t. j. byłby 7,7 razy mniejszy, a więc moglibyśmy wnioskować, że współzależność omawianych cech istnieje w zbiorowości generalnej.

Jednak spółczynnik współzależności posiada tę ujemną właściwość, że wyraża ściśle współzależność między zmiennemi tylko w wypadku regresji linjowej. Jeśli istnieje między zmiennemi związek funkcyjny krzywolinijny, to współczynnik r, mimo najściślejszej zależności tych zmiennych, może się zbliżyć do zera, może się nawet równać zeru. Przez tę właściwość spółczynnika r zwęża się jego znaczenie. Dlatego to, kiedy rozpatrujemy stosunki zależności, a nie jesteśmy pewni, czy regresja jest linjowa, należy z wielką ostrożnością wyjaśniać wartość spółczynnika r i możliwie uzupełniać ją przez obliczenie innych liczb. Paerson radzi wprowadzać t. zw. stosunek współzależnościowy (η)

$$\eta = \frac{SAy}{y}$$

który jest stosunkiem odchylenia średniego średnich arytmetycznych kolumn, czy rzędów w tablicy współzależności od odchylenia średniego przypadającej zmiennej. Stosunek współzależnościowy jest niezależny od zmienności kierunków współzależności między zmiennemi. W naszym przykładzie

$$\eta = 0{,}46\,,$$

a z błędem średnim, który się oblicza według wzoru

$$\sigma_\eta = \frac{1 - \eta^2}{\sqrt{n}}$$

stosunek współzależnościowy

$$\eta = 0{,}46 \pm 0{,}13.$$

Ze stosunku współzależnościowego spostrzegamy, że współzależność jest już dość wyraźna, a biorąc pod uwagę znacznie mniejszy spółczynnik współzależności, tworzymy hipotezę, że może tu istnieje związek funkcyjny krzywolinijny. Związek funkcyjny krzywolinijny między zmiennemi wśród zjawisk gospodarskich może się łatwo zdarzyć w praktyce gospodarczej. Mamy do czynienia z regresją krzywolinijną przedewszystkiem w tych wszystkich wypadkach, w których wzrost intensywności od pewnej granicy poczyna wyraźnie powodować spadek dochodów czystych. Może się również zdarzyć w jakichś warunkach ekonomicznych, że największe dochody czyste dają gospodarstwa najbardziej ekstensywne i najbardziej intensywne, albo gospodarstwa małoinwentarzowe i bardzo silnie nasycone inwentarzem dochodowym. Wtedy linja zależności będzie krzywa, a spółczynnik r nie będzie trafnie wyrażał tej współzależności.

Celem przekonania się, czy hipotetycznie postawiona regresja krzywolinijna istnieje w zbiorowości generalnej, obliczamy t. zw. wskaźnik krzywizny regresji, który się oblicza według wzoru

$$I = \eta^2 - r^2$$

Jeśli wskaźnik jest małą liczbą, nie przewyższającą znacznie trzykrotnej wartości swego błędu średniego, obliczonego według wzoru

$$\sigma_I = 2\sqrt{\frac{I}{n}}$$

nie mamy podstaw do twierdzenia, że w zbiorowości generalnej istnieje regresja krzywolinijna.

W naszym przykładzie wskaźnik krzywizny = 0,09, a jego błąd średni = 0,10, a więc wskaźnik jest mniejszy od swego błędu. Nie mamy żadnych podstaw do wnioskowania, że kierunek regresji w zbiorowości generalnej jest krzywolinijny. Gdybyśmy zaś przyjęli, jak poprzednio, że przy tej samej tendencji współzależności w zbiorowości cząstkowej, zbiorowość ta zawierałaby 400 spostrzeń, w takim razie błąd średni wskaźnika krzywizny równałby się 0,03, a więc byłby trzy razy mniejszy od wskaźnika. W takim wypadku i to tylko przy założeniu, że zdarzenia rozpatrywane ulegają prawu Gauss'a, moglibyśmy już z dość dużem prawdopodobieństwem wnioskować, że dalsze zwiększanie uprawy buraka cukrowego, średnio biorąc, nie opłaca się w danym okręgu. Praktycznie należy przyjąć, że wskaźnik krzywizny regresji powinien być kilkakrotnie razy większy od swego błędu, ażebyśmy mogli wnioskować, że w zbiorowości generalnej panuje regresja krzywolinijna.

IV.

To, że zjawiska natury rolniczej ulegają prawu Gauss'a w wyjątkowych wypadkach, osłabia znaczenie teorji błędów w analizie tychże zjawisk. Jednak badacze posługują się z dobremi wynikami teorją błędów w nauce o produkcji roślinnej i zwierzęcej. Wszystko przemawia za tem, że rachunek prawdopodobieństwa da się z powodzeniem zastosować również do zdarzeń ekonomicznych w gospodarstwach wiejskich.

Reasumując dotychczasowe uwagi i powołując się na przytoczone przykłady, dochodzimy do takich wniosków.

O ile to tylko jest rzeczą możliwą, wydawnictwa statystyczne, dotyczące ekonomiki gospodarstw wiejskich powinny zawierać surowy materjał, ażeby każdy z badaczy miał możność dowolnego opracowania danych według swego planu. Kiedy zaś badacz podaje wyniki swych opracowań, powinien zachować następujące przepisy, jako minimum wymagań:

1. Średnia arytmetyczna zbiorowości cząstkowej powinna być podana ze swym błędem według wzoru:

$$\sigma_A = \frac{\sigma}{\sqrt{n-1}}$$

jeśli liczebność zdarzeń w zbiorowości generalnej jest nieznana, a według wzoru:

$$\sigma_A = \sqrt{\frac{N-n}{N(n-1)}} \cdot \sigma$$

jeśli liczebność zdarzeń w zbiorowości generalnej może być oznaczona.

2. Przy porównywaniu dwóch średnich arytmetycznych należy podać błąd średni różnicy między niemi według wzoru

$$\sigma_{A_1 - A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2}},$$

jeśli między badaniami zmiennemi wykluczona jest wszelka zależność, a według wzoru

$$\sigma_{A_1 - A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2} - 2\,r\,\sigma_{A_1} \cdot \sigma_{A_2}}$$

jeśli zależność jest domniemana.

3. Badanie współzależności między zmiennemi sposobem uproszczonym za pomocą szeregów uporządkowanych według wielkości jednej cechy, jak się to robi powszechnie w literaturze, jest zawsze pożyteczne dla analizy statystycznej, gdyż pozwala badaczowi wnikać w wiele ciekawych szczegółów. Jeszcze większe bogactwo spostrzeżeń daje inny, równie prosty sposób analizy, polegający na przeprowadzeniu linji pojedyńczej przez

szereg punktów, które w układzie dwóch współrzędnych są skojarzeniem dla każdego z kolei osobnika dwóch jego cech. Jednak badacz powinien zawsze podać spółczynnik współzależności r z jego błędem średnim według wzoru

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}$$

4. O ile uproszczone badania wykazują możliwość istnienia regresji krzywolinijnej w zbiorowości generalnej, należy obliczyć i podać stosunek współzależnościowy — z jego błędem średnim według wzoru

$$\sigma_\eta = \frac{1 - n^2}{\sqrt{n}}$$

i obliczyć wskaźnik krzywizny regresji z jego błędem średnim według wzoru

$$\sigma_I = 2\sqrt{\frac{I}{n}},$$

ażeby tym sposobem dać możność czytelnikowi krytycznej oceny stosunków panujących w owych zbiorowościach.

Wprowadzenie w ten sposób rachunku prawdopodobieństwa do ekonomiki gospodarstw wiejskich jako wykładnika liczbowego zdrowego sensu, może mieć duże znaczenie dla praktyki rolniczej, bo wskaże rolnikowi, z jak wielkiemi pomyłkami ma on do czynienia, kiedy chce uogólniać swe nieliczne spostrzeżenia. Rachunek prawdopodobieństwa uczy rolnika ostrożności w kierowaniu planami gospodarczemi.

Stefan Moszczeński.

Sur l'unification des méthodes de l'élaboration statistique des données de comptabilité agricole.

Institut d'Economie Rurale de l'Ecole Supérieure d'Agriculture à Warszawa.

Résumé.

L'application de la théorie des erreurs, à l'analyse des phénomènes dans l'agriculture, ne peut donner des résultats exacts, car les phénomènes sus-dits, dans des cas exceptionnels obeissent à la loi de Gauss. Cependant l'investigateur applique la théorie des erreurs à l'élèvage des animaux, à la culture des plantes, avec de bons résultats. Il est déjà bien temps, que, pour l'économie rurale, le calcul des prohabilités, ait pleinement droit de cité.

Résumant nos remarques et en nous reférant aux exemples donnés, nous arrivons aux conclusions suivantes.

Autant, que la chose est possible, les publications statistiques, traitant de l'économie rurale, devraient comprendre un matériel brut, pour que chaque investigateur ait la possibilité d'élaborer librement, selon son propre plan, les données réunies ainsi. Quand l'investigateur présente les résultats de son travail, il doit s'en tenir aux préceptes suivants, étant le minimum de ce qui est exigeable.

1. La moyenne arithmétique de la population partielle doit être présentée avec son erreur, selon le modèle:

$$\sigma_A = \frac{\sigma}{\sqrt{n-1}}$$

si le chiffre des événements dans la population générale est inconnu, et selon le modèle

$$\sigma_A = \sqrt{\frac{N-n}{N(n-1)}} \cdot \sigma,$$

si le chiffre des événements dans la population générale est connu.

2. En comparant deux moyennes arithmétiques il faut présenter l'erreur moyenne de la différence entre elles, selon le modèle:

$$\sigma_{A_1-A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2}}$$

entre les deux variables étudiées, toute dépendance est exclue, et selon le modèle:

$$\sigma_{A_1-A_2} = \sqrt{\sigma^2_{A_1} + \sigma^2_{A_2} - 2\, r\, \sigma_{A_1} \cdot \sigma_{A_2}}$$

si la corrélation est supposée.

3. Les études de la corrélation entre variables, faites d'une manière simplifiée, à l'aide de séries ordonnées selon la grandeur d'une caractéristique, comme cela se fait généralement dans la litérature, sont toujours utiles pour l'analyse statistique, car elles permettront à l'investigateur d'approfondir beaucoup de de détails curieux. Un autre simple genre d'analyse donne aussi beaucoup d'observations. Il consiste dans le tracement d'une ligne unique entre les séries de points, qui pour une coordonnée rectiligne plane, caractérisent l'union des points de chaque particulier à tour de rôle. Cependant l'investigateur doit toujours donner le coéfficient de corrélation (r) avec son erreur moyenne, selon le modèle:

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}$$

4. Si les investigations simplifiées prouvent la possibilité de la regression curviligne dans la population générale, il faut calculer et donner le rapport de corrélation (η) avec son erreur moyenne selon le modèle:

$$\sigma_\eta = \frac{1-\eta^2}{\sqrt{n}},$$

et calculer l'index de curbure de la regression ($I = \eta^2 - \eta^2$) avec son erreur moyenne selon la modèle:

$$\sigma_I = 2\sqrt{\frac{I}{n}}$$

pour permettre de cette façon au lecteur, d'analyser critiquement les conditions, qui régnent dans ces populations. L'introduction du calcul des probabilités dans l'économie, rurale comme exposition numérique du sens commun, a une importance toute spéciale pour l'agriculture, car elle démontre à l'agriculteur les grandes erreurs, qu'il commet, quand il veut génèraliser ses observations restreintes. Le calcul des probabilités enseigne à l'agriculteur la prudence dans la direction de ses plans agricoles.

